Northern Economic Review
ISSN: 0262-0383
Vol 14, Issue 2 (2023)
https://nerj.org/

Knowledge Discovery Using Web Mining Technique

D.Saravanan, Dr. Dennis Joseph Dr. KVSSN Murty
Faculty of Operations & IT
ICFAI Business School (IBS), The ICFAI Foundation for Higher Education (IFHE)
(Deemed to be university u/s 3 of the UGC Act 1956)
Hyderabad-India.

Abstract

Data Mining, provided the rules are strictly utilitarian terms, decreased in size large. Among the set of items in a transaction connotation instruction withdrawal, knowledge discovery in databases is considered as essential responsibilities. Overcome this drawback, many times more abstract representations of the item set, redundancy reduction, and post-processing proposed in literature. However, on the basis of statistical information to the public, this technique does not work in all types of input sets; few will give better results, and few will give average outputs only. Thus, in order to overcome the burden the amount of instructions in a more efficient post-processing step is important to help the decision maker. In the proposed methods a new extraction approach is proposed to find the association among the given rules planned.. First, we have in mind in order to improve the processing task is to propose the integration of user knowledge. Second, propose user expectations peraventumenak Schema specification language to extend the rule. Also, an interactive framework to examine the entire procedure is designed to assist the user. The largest fixed of rules for the use of our new approach, it will decrease the total amount of instructions drastically according to the processing by integrating expert knowledge of the domain, complete. Also, the page ranking algorithm, the necessary data is mined.

Key Terms - Information Detection, Data Extraction, Information withdrawal, Raking, Rule formulation, Source extraction, network connectivity.

I. INTRODUCTION

Today, people on the web serve as a valuable source of information. Data on the internet comprises tiny, indivisible units within the system, and web pages are typically considered as a single semantic representation. However, when viewed as a whole, a web page may contain various elements, such as navigation, decorations, and interaction material, integrated with different types of content. This source contains a diversity of information, not all of which is necessarily relevant. To address this issue, systems are designed to detect and extract the content of web pages. The semantic web plays a crucial character in attractive the output of source extraction. This process aims to improve the organization and structure of information available on web pages. To mitigate the drawbacks of web information, many researchers employ content-based searching or create indexing terms to extract the relevant web pages. Additionally, they utilize the structure of web pages to locate the desired content. In all of these techniques, users must develop separate terminologies to effectively extract the information they need..

Existing System

In the proposed technique, information extraction is carried out using a rule-based method. Initially, rules are defined among the datasets, and these initial rules serve as the basis for further information extraction. These rules are then tested with specific outcomes in mind, and the information should meet certain predefined criteria. If the outcomes do not align with the expected results, the rules may need to be redefined, and new sets of rules are created and tested for effectiveness. The acceptance of these rules is based on the outcomes they produce. This process can be time-consuming, especially when defining rules based on complex and extensive input sets. Each time a user specifies new rules based on the initial sets, unexpected outcomes may occur, which can create additional work for researchers. Both expected and unexpected outcomes are considered when evaluating the quality of the datasets. Based on these

outcomes, a final set of rules is defined. To enhance understanding and facilitate further operations, the final outcomes are often presented in graphical form.

Disadvantage of existing techniques

1. The formation of initial rule sets is time-consuming.
2. Extracting sub-cluster rules can be challenging, especially after the initial rule formation.
3. The process of rule formation hang on on the participation sets.
4. Often, the proposed rules result in unexpected and incorrect outcomes.
5. Rules are consistently accepted based on predetermined threshold values.
6. Measuring these values is a complex task.

Proposed System

The primary limitation of the existing system lies in the complexity of creating the initial set of rules and establishing similarity among defined rules. In the proposed system, this challenge can be mitigated through the use of a rule-purification approach. This method aims to reduce unexpected outcomes and exclusions in various analyses. By confirming the formation of rules, it ensures that users can achieve more predictable results. Additionally, it enhances the overall performance of rule formation, allowing users to save
time by eliminating the need for initial rule creation and any subsequent rule pruning.

Advantages of proposed systems.

1. Initial rule formation time is reduced.
2. Creating sub-cluster rules becomes easier and faster.
3. Unexpected result formations are minimized.
4. Users can obtain rules more quickly and accurately.
5. Rule formation is streamlined.
6. The generation of interesting rules is enhanced.

II Network extraction

Data extraction, or knowledge extraction, is the process of retrieving essential content from a stored databank. Users need to specify key terms or search queries to extract the necessary information, which can include documents, data, images, and more. Based on the search terms provided, the relevant information is extracted and delivered to the user. When extracting information from text, it is referred to as "text extraction" or "text mining." Similarly, when extracting information from a network, it is known as "network mining," where users can extract patterns or network pages.

Network structure extraction

Network structure extraction operates on the principles of graph methodology. It involves extracting nodes and understanding how they are interconnected based on the provided input information. This categorizes network structure extraction into two main types.

1. Retrieving the connection pattern on the network.

2. Mining the document structure:

Web mining is the process enchanting the different domain peoples either in private sector or in public sector based on this equipment has many advantages. More Earth at the end of the technology, the result is personalized marketing, ecommerce enabled to do. Government agencies to classify threats and the technology of the war against terrorism. The application is capable of predicting the benefit of the community by identifying criminal activities. Clients are able to establish a good relationship with the companies by attracting their requirements. Corporations can better appreciate customer requests and customer demand for faster, they can act. The unique nature of the concerns when used for data mining the Web, by itself, create problems, but this technology. Web mining ethical issue involving the invasion of privacy is the most criticized. Privacy is a unique use of information, or the spread of the data obtained from this study will be made without their knowledge or consent occurs in particular, lost, and is considered to be clustered to form profiles; There are no personal details

that data package, before being anonymous. Judging by the users of these applications to de-reference them with mouse clicks. De-individualization, and the characteristics of the study on the basis of merit rather than on the basis of their own individual characteristics, and the population is defined as a course of treatment As controversial as it is tough to classify the characteristics of the applications, and the use of such methods with attributes such as strong against the attribute.

III. EXPERIMENTAL SETUP

3.1 Input from the sender.

3.2 Construction Province outlook.

3.3 Creating the success using the deters ion techniques.

3.4 Applying the proposed procedure.

3.1 Input from the sender.

Information is entered with the user's assistance. Before a user can input data, they must first be authorized. This authorization process validates the user's access and helps prevent unauthorized entries into the network, thereby reducing various complications. Both the sender and the receiver must be approved by the system, ensuring secure transactions within the community. Anytime the sender's credentials are validated, incorrect credentials trigger warnings or restrict permission for entry, promoting alertness in the system. Server-side monitoring allows administrators to identify the number of attempts made by a particular user, keeping a record of unauthorized access attempts and related information. This valuable data aids administrators in enhancing security applications. If the input credentials are correct, the system recognizes the user as legitimate, approves access, and provides access to information. The system also enables the creation of various privileges among users, allowing certain users to create, read, and modify privileges while others are granted read-only access. The assignment of these privileges is determined by the administrator based on the type of authorization each user should have.

3.2 Construction Province outlook.

The creation of a province outlook is one of the most crucial tasks in network mining. In the data extraction process, users typically input key terms or specific query terms. Based on these terms, information or specific content is extracted and sent back to the user. At the time of input, a corresponding link is sent to the database bank. Any information stored in this database has a corresponding link created, which helps retrieve the necessary content. Subsequently, an index is generated for all stored content, significantly improving search times. Although this process initially requires more time, it ultimately reduces search complexity, allowing users to access more relevant datasets quickly. A ranking system is established for each search based on the user's input, and the results are accumulated and stored. Documents with higher ranking values are presented first. This entire process is depicted in Figure 1.

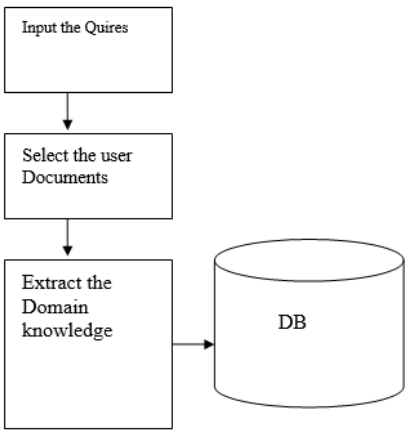

Fig 1. Creation of Users information

3.3 Creating the success using the detersion techniques.

Initially, rules are defined among the datasets, and these initial rules serve as the foundation for further information extraction. These rules are then tested with specific outcomes in mind, and the information should meet certain predefined criteria. If the outcomes do not align with the expected results, the rules may need to be redefined, and new sets of rules are created and tested for effectiveness. Each time a user needs to specify new sets of initial rules after testing the goodness, the second and third stages of inter-cluster rules are established. This process becomes time-consuming

when dealing with complex and extensive datasets, this process is shown in the figure 2 to 6. To address this issue, a filtering technique is proposed. It checks each input rule for any unwanted rules that may have been extracted in the initial phase itself. This helps avoid the need to test the goodness of the result.

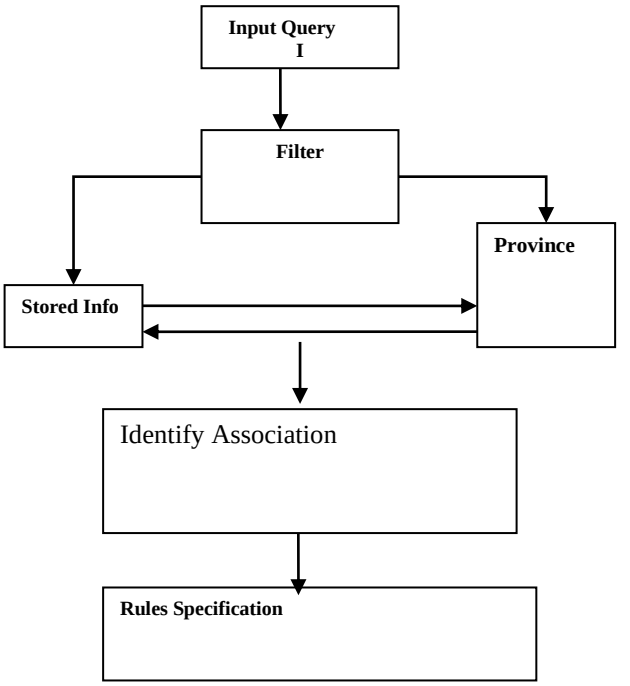

Fig 2Work function of Filiter

3.5Applying the proposed procedure.

A ranking system is established for each search based on the user's input, and the results are accumulated and stored. Documents with higher ranking values are presented first . Any information stored in this database has a corresponding link created, which helps retrieve the necessary content. Subsequently, an index is generated for all stored content, significantly improving search times. Each time a user specifies new rules based on the initial sets, unexpected outcomes may occur, which can create additional work for researchers. Both expected and unexpected outcomes are considered when evaluating the quality of the datasets.

IV. Conclusion and Future enhancement

Discovered a large amount of instructions, this work discourses a fresh extraction method is proposed to find the association among the given rules planned. The main advantage of the proposed work is stated below. Ontology schemas and the rule: First, we use two different types of relationship instruction withdrawal peraventumenak suggest to participate the user's knowledge. On the one hand, the later-processing stage in the study of the field of the user province knowledge in the field of information integrity. A specification language, we expanded the rule that a new procedure Schemas, propose. Discovered, in particular with regard to the rules offered by the researcher's anticipations and aims, it shown in the fig 7. Second, the rule applies to the operators on schemas, for a moment, in directive to monitor the user through the proposed data processing step. Thus, many of the varieties of movements that conditioning and filtration, the user will get. Finally, our ARIPSO interactive framework, the rule depends on mining operators, analysis, and filtering rules work is supported throughout the process the user is urged interesting rules / she allows an easy selection. By using our new approach on a large database of the questionnaire, we have several dozen or fewer rules in imperative to decrease the amount of later-processing stage, allowing the combination of skilled information of the domain. Also, throughout the process of interactive expert to verify the quality of the filter rules. Privileges and managed using an expansion queries.

VI . EXPERIMENTAL RESULTS

Fig 3. Data extraction screen

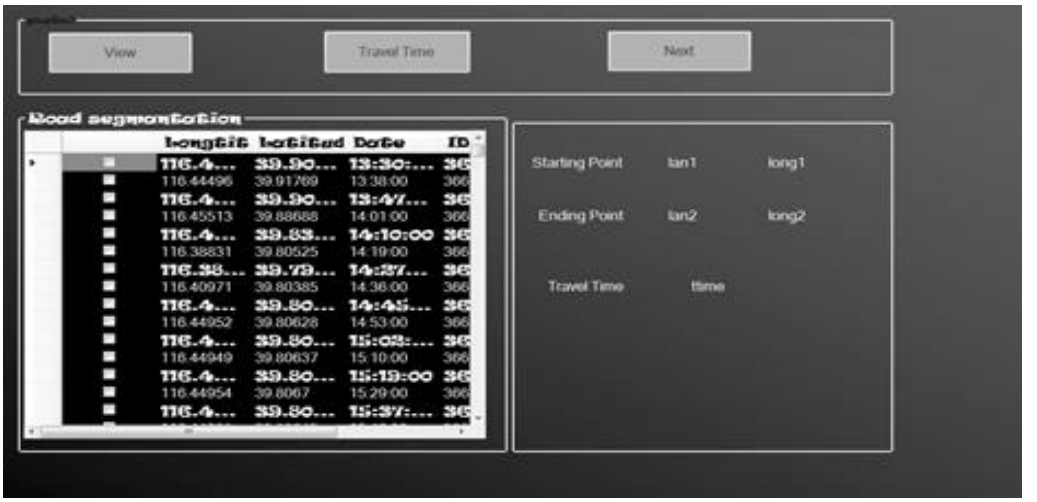

Fig 4. Network link collected and stroed.

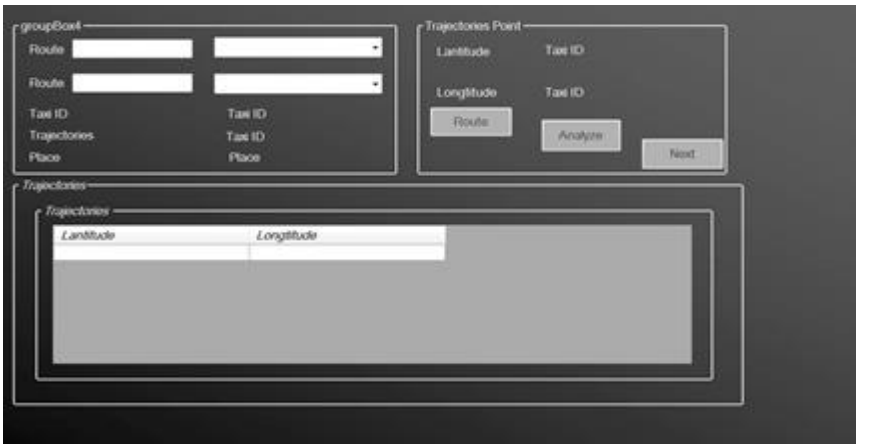

Fig 5. Network path confirmation.

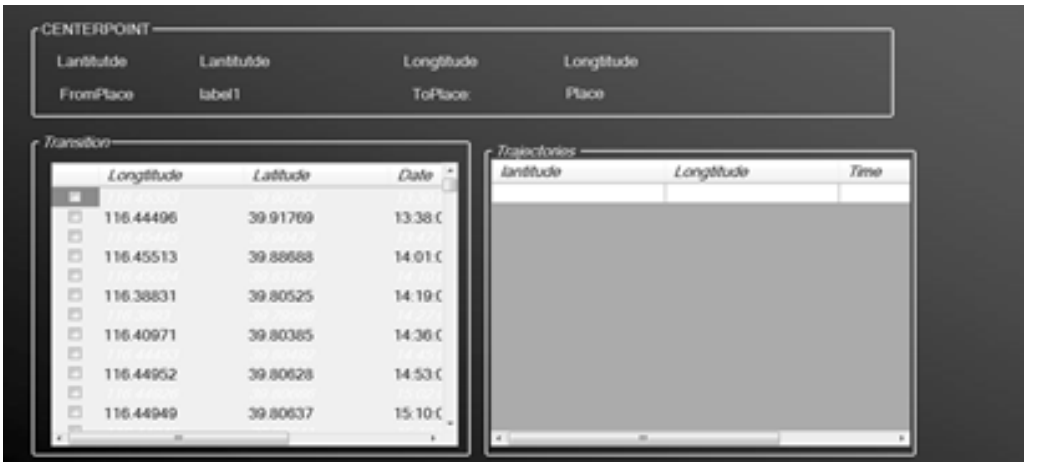

Fig 6 Network link path and resource information's.

Fig 7.Network searching results.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD, pp. 207-216, 1993.

[2] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996.

[3] A. Silberschatz and A. Tuzhilin, "What Makes Patterns Interesting in Knowledge Discovery Systems," IEEE Trans. Knowledge and Data Eng. vol. 8, no. 6, pp. 970-974, Dec. 1996.

[4]D.Saravanan V.Soma sundaram "MATRIX BASED SEQUENTIAL INDEXING TECHNIQUE FOR VIDEO DATA MINING Journal of Theoretical and Applied Information Technology 30th September 2014. Vol. 67 No.3.

[5] M.J. Zaki and M. Ogihara, "Theoretical Foundations of Association Rules," Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98), pp. 1-8, June 1998.

[6] D.Saravanan, Dr.S.Srinivasan (2010).Indexing ad Accessing Video Frames by Histogram Approach, In the Proc. Of International Conference on RSTSCC 2010, 196-1999.

[7] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemset Algorithm," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.

[8] D.Saravanan, Dr.S.Srinivasan(2013) "Video information retrieval using :CHEMELEON Clustering" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume-02,Issue 01, January –February 2013, Pages 166-170.

[9] D.Saravanan, A.Ramesh Kumar, "ContentBased Image Retrieval using Color Histogram", International journal of computer science and information technology (IJCSIT), Volume 4(2), 2013, Pages 242-245, ISSN: 0975-9646.

[10] J. Li, "On Optimal Rule Discovery," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 460-471, Apr. 2006.

[11] D.saravanan, Dr.S.Srinivasan (2012). Video image retrieval using data mining Techniques, Journal of computer applications (JCA), Vol V,Issue 01, 2012. 39-42.

[12] M.J. Zaki, "Generating Non-Redundant Association Rules," Proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 34-43, 2000.

[13] D.Saravanan, Dr.S.Srinivasan, (2013). , Matrix Based Indexing Technique for video data, Journal of computer science, 9(5), 2013, 534-542.

[14] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Efficient Mining of Association Rules Using Closed Itemset Lattices," Information Systems, vol. 24, pp. 25-46, 1999.

[15] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hatonen, and H. Mannila, "Pruning and Grouping of Discovered Association Rules," Proc. ECML-95 Workshop Statistics, Machine Learning, and Knowledge Discovery in Databases, pp. 47-52, 1995.